Reducing Cycle Time Has Many Benefits

Doug Sutherland – KLA-Tencor Corporation

Metrology and inspection steps usually account for about 5% of a wafer fab's total cycle time, but the value they provide in terms of improved yield is typically an order of magnitude greater than the cycle time cost that they impart to the process. However, a cycle time management program, to be successful, must be a fab-wide activity with equal attention paid to reducing the cycle time at every toolset in the fab. In recent years there has been a trend in wafer fabs away from maximizing tool utilization, which reduces cost per wafer, and toward minimizing cycle time, which increases revenue and profit. The two objectives are at odds with each other, as decreasing utilization decreases cycle time but also decreases producti-vity. The optimal operating point is one that strikes a balance between the two.

Reduced cycle time (CT) has many benefits; the primary one is faster time to market. The price of nearly all semiconductor products, DRAM, Flash, Logic, etc., declines rapidly over time - typically from 50% to 80% per year from the time the product is first released. Shorter CT ensures less decline in price from the time the product enters manufacturing to the time it reaches the market, thereby commanding a higher ASP. The other benefits revolve around having shorter cycles of learning (COL) and reduced work in progress (WIP). In R&D shorter COL equates to shorter development time and, when transferred to production, faster yield ramp. The relationship between WIP and CT is expressed through what is called Little's Law:[1]

WIP = (CT) x (Start Rate)

From the equation above it can be seen that, for a given start rate, the WIP will decrease linearly with CT. The advantage of this are that there are fewer lots in the fab at any given time, which reduces overhead, exposes fewer lots to any required process changes, and reduces the number of lots at risk during any yield excursions that may occur. Carrying less WIP also means there are fewer unfinished goods on hand when the market turns down. One of the best value statements for cycle time was summarized by Clayton Christensen[2] who said,

"Extending development an extra day, to get a stepper or process qualified, is like paying $3.44 for every wafer that the factory will make. In addition, if it takes one more day to reach mature die yield, it is like paying $1.35 for every wafer that will be made, or if the cycle time is one day longer, it is like paying $3.04 per wafer."

From this quote we can get a feel for the value of CT, which is approxi-mately $1 million per year for every day of CT reduction (30,000 WSPM x 12 months x 3.04 per wafer = $1.1 million per year).

Mathematically, CT is equal to the queue time (the time a lot spends waiting to be processed) plus the processing time (the time it spends in the tool). The processing time is a straight forward calculation but the queue time (QT) is the product of three separate functions.[3]

QT = {f(Variability)} {f(Utilization)} {f(Availability)}

There is no single *correct* version of the equation above; it comes in several incarnations with varying degrees of complexity depending on the level of detail one wishes to incorporate. However, essentially all the mathematical expressions of QT have the following four features in common:

1) A system with no variability has no queue time: when f(Variability)=0, QT=0
2) f(Utilization) is proportional to 1/(1-Utilization): CT increases exponentially with increased utilization.
3) f(Utilization) is also proportional to 1/(# of tools): CT decreases with more tools.
4) f(Availability) is proportional to 1/(Availability): CT decreases exponentially with increased availability (uptime).

Low utilization and high availability do not in themselves ensure that the CT will be low; only by eliminating all sources of variability can one guarantee that the QT will vanish. Mathematically, variability is measured as the standard deviation of a system divided by its average. In a wafer fab variability comes from three main sources:

1) Variability in the lot arrival rate
2) Variability in the lot processing time
3) Variability in the downtime of the tool

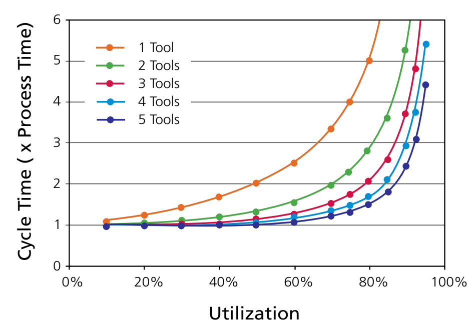

Figure 1: Cycle Time vs. Utilization for a toolset with 1 to 5 tools. Units of CT are in multiples of the tool's processing time. Large fabs with more tools in each toolset have an advantage because they can run at higher utilizations without as much impact on CT.

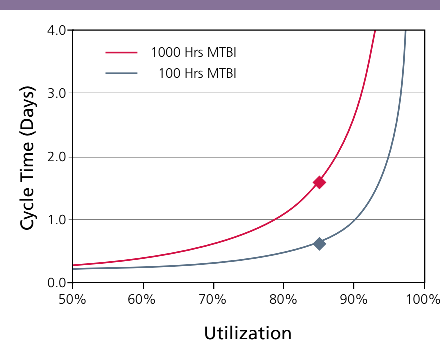

Figure 2: Cycle Time vs. Utilization for two toolsets with the same availability (95%) but different MTBI (and MTTR). From a CT perspective, *for the same availability*, it is better to have many short down events (MTBI = 100 Hrs) than comparatively fewer long ones (MTBI = 1000 Hrs). The difference in this example is about 1 day at 85% utilization.

Figure 1 shows the operating curve (a plot of CT vs. utilization) for identical toolsets with one to five tools assuming unit varia-bility and 100% availability. The salient point here is that changing from one tool to two does much more than simply double the capacity. For the same CT as one tool at 60% utilization, you can run two tools at nearly 80% utilization. Not only do you have twice as many tools but each one of them is processing about 30% more wafers — a 260% improvement. The impact of having n+1 tools is less dramatic with larger toolsets but the same principle applies and this is one of the underlying tenets of the economy of scale enjoyed by larger wafer fabs. Large fabs generally have lower CT *and* lower cost per wafer because they can run their tools at higher utilization without climbing into the steepest part of the operating curve.

In addition to the number of tools, the other first-order effects on CT are the related variables, availability and utilization. If we artificially set availability and utilization to 95% and 85%, respectively, we can see some interesting and unexpected trends in cycle time. For instance, for a given availability, CT actually *increases* with increasing MTBI. That is, it is better to divide the same downtime into many short events (low MTBI and low MTTR) rather than fewer long events (high MTBI and high MTTR), as shown in Figure 2. This isn't a problem in terms of tool design, as we usually make the assumption that higher MTBI (fewer system-down events) equates to higher availability. However, from a service perspective we often adopt the pragmatic philosophy that "if we're here to fix Problem A we might as well do adjustments B, C and D at the same time." This well-intentioned approach increases the tools MTBI and MTTR but does not substantially improve the availability (i.e., the total repair time remains unchanged) and consequently increases the CT. Often our best intentions tend to be counter-intuitive (and counter-productive) in terms of reducing our customer's CT.

Another interesting case is that of matching, as having dedicated (or "golden") tools is one of the worst things for CT. Figure 3 shows the impact to CT of having five matched brightfield inspection tools inspecting five layers in the process versus 4 matched tools inspecting four layers and one dedicated (golden) tool inspecting one layer (for the sake of simplification this assumes 100% sampling). Instead of having five layers all experiencing cycle times represented by the operating curve for five tools (see Figure 1) you have four layers with cycle times represented by the curve for four tools and one layer with a CT represented by the curve for a toolset with only a single tool. The net effect of unmatched tools in this case is to double the total CT for that toolset (Figure 3). Fabs can mitigate the effect of this by treating the tools as if they were matched whenever the golden tool is unavailable (i.e., instead of holding the lot to wait for the golden tool they run it on one of the other tools), but this comes at the cost of increased beta risk.

Service contracts provide a threefold advantage for cycle time management. First and foremost, they increase the availability of the tool. This in turn has the added advantage of automatically reducing the utilization (utilization is equal to production time divided by available time). Finally, service contracts significantly reduce the variability in the downtime, which itself is a significant contribution to CT. Figure 4 shows the operating curves for six brightfield inspection tools under two different conditions: one where their reliability characteristics are typical of billable tools and the other where the six tools are covered under a service contract. The faster response time (less time down, higher availability) and the reduced variability in the downtime result in a CT reduction of about 1.9 days. Another key factor is that because utilization

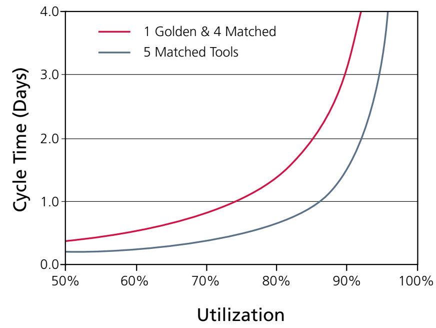

Figure 3: Dedicating layers to specific "golden tools" creates a "single tool" environment (see Figure 1) and causes a dramatic increase in cycle time that is exacerbated at higher utilizations. In this case the cycle time almost doubles as a result of having unmatched tools.

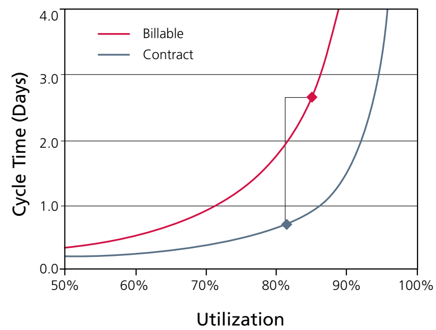

Figure 4: The cycle time impact of converting six brightfield tools from billable to service contract. The increased availability achieved by having the tools on contract flattens out the operating curve and also contributes to lower utilization (utilization equals production time divided by available time). For tools that are at 85% utilization when billable, a service contract can reduce the cycle time by 1.9 days.

is equal to the production time divided by the available time, a toolset running at 85% utilization while billable can be run at about 82% utilization under service contract.

As the IC industry becomes increasingly driven by consumer electronics, cycle time (or equivalently, time to market) will become ever more important to wafer fabs as they strive to produce exactly the right amount of product at exactly the right time. Being the first IC manufacturer to provide engineering samples to a prospective customer can result in design wins that could literally make or break its business. Similarly, being caught with hundreds of millions of dollars worth of WIP still in the pipeline when the market enters a downturn or when the consumer simply moves on to "the next new thing" can make the difference between a year that closes with a profit and a year that closes with a loss. There are, and will continue to be, niches within the IC industry where cycle time is less important, but the general trend for the foreseeable future is in the direction of decreased cycle time and increased operational efficiency. As a result of this, more emphasis will be given to products and services that provide higher availability and that reduce variability in the wafer fab environment.

References

1. W.J. Hopp and M.L. Spearman, McGraw-Hill, "Factory Physics", 2001, p. 223.
2. Clayton Christensen, "Solid State Technology", August 2001.
3. W.J. Hopp and M.L. Spearman, McGraw-Hill, "Factory Physics", 2001, p. 325.